

Phase 3 — Handling Data Products from ESO Public Surveys, Large Programmes and Other Contributions

Magda Arnaboldi¹
 Jörg Retzlaff¹
 Remco Slijkhuis¹
 Vincenzo Forchì¹
 Paulo Nunes¹
 Diego Sforza¹
 Stefano Zampieri¹
 Thomas Bierwirth¹
 Fernando Comerón¹
 Michèle Péron¹
 Martino Romaniello¹
 Dieter Suchar¹

¹ ESO

Phase 3 represents the final step in the execution of ESO large programmes and public surveys, which starts with the submission of the letters of intent for public surveys and proposals for large programmes, i.e. Phase 1, and continues with the preparation and submission of observing blocks for service mode observations, i.e. Phase 2. In this paper we present the new Phase 3 infrastructure deployed on 10 March 2011. This infrastructure supports the reception, validation and publication of data products from the public survey projects and large programmes to the ESO Science Archive Facility.

[Access to the survey data products by the astronomical community](#)

The ESO public survey projects on the near-infrared 4-metre VLT Infrared Survey Telescope for Astronomy (VISTA) (Emerson et al., 2006) and the optical 2.6-metre VLT Survey Telescope (VST) (Capaccioli et al., 2005) are ambitious projects that range from very wide area surveys with short exposures, like the VISTA Hemisphere Survey (VHS), which aims at covering the whole southern hemisphere, to deep surveys concentrating on small areas or even a single pointing on the sky, and going very deep. Typical examples of the latter are the UltraVISTA and VIDEO surveys (for an overview of the six VISTA and the three VST public surveys see Arnaboldi et al. [2007]). A plot reproducing the survey areas of the six VISTA surveys is shown in Figure 1. In addition to the imaging surveys, ESO opened a call for spectro-

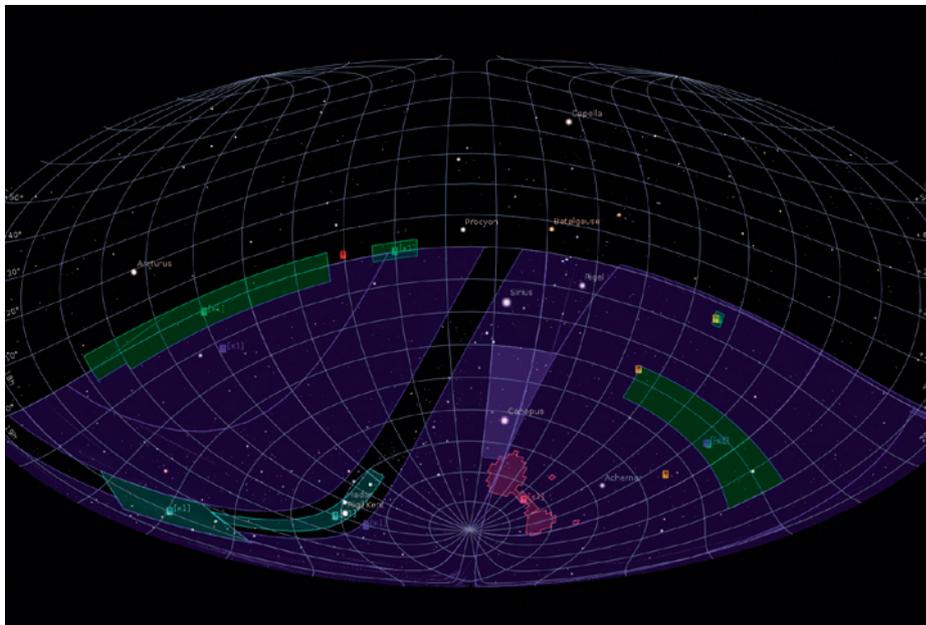


Figure 1. Summary of the regions of the sky covered by the six VISTA surveys: VHS is shown as the blue shaded areas; VIKING areas in light green; VVV areas in light blue; VMC areas in purple; and the ULTRA-VISTA and VIDEO fields in red and yellow, respectively.

scopic public surveys and the selected projects will soon follow the same policies and procedures for telescope operations, access to data products and publication.

The raw data collected at the survey telescopes for the different projects amounts to about 700 GB per month, which in turn are condensed into a few terabytes of data products each year. Because of the legacy value of the public survey projects, ESO's policy is to ensure their long-term archival value by supporting easy access to the data products and fostering their wide scientific use by the astronomical community at large, beyond those projects initially identified by the survey teams.

Phase 3 policies and concepts

Policies

As stated in the ESO Council document on the VLT/VLTI science operation policies (Meeting#104, 17–18 December 2004), the ESO Science Archive Facility (SAF) is the collection point for the survey products and the primary point of

publication/availability of these products to the ESO community. As part of the implementation of the Council's recommendations, a dedicated group within the Data Products Department (DPD) was set up to oversee the definition of the requirements for Phase 3 and their implementation for the validation, ingestion and publication of data products from the various ESO projects in the SAF. The Phase 3 tools, user manuals and definitions of data standards are all available on the ESO website¹.

The dependencies between telescope time allocation and the delivery of data products to the ESO SAF is made explicit in the policies for public surveys. In fact, further allocation of observing time at the survey telescopes beyond the first year and a half, and the scientific follow-up at the VLT of the public survey targets is subject to timely delivery of the survey products and their compliance to the specifications detailed during Phase 1. The Public Survey Panel that initially took part in selecting the public survey projects will periodically review the progress of the surveys and report to the OPC for any additional allocation of telescope time to these projects.

Concepts

The Phase 3 concepts provide the framework that supports the data submission process and facilitates data access

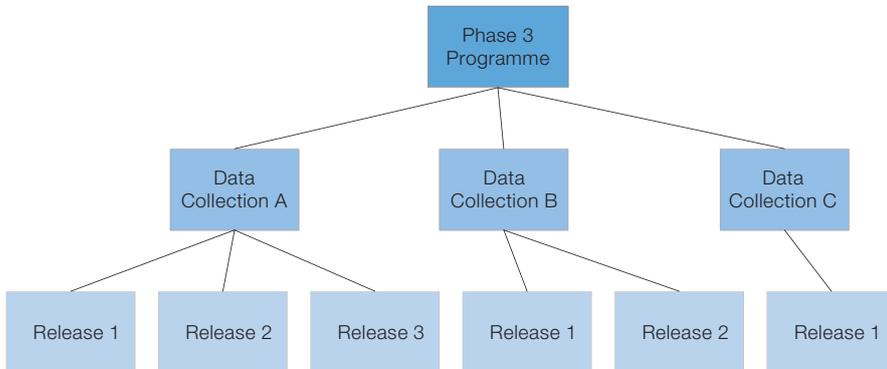


Figure 2. Block diagram illustrating the hierarchy of the Phase 3 concepts: from Phase 3 programme to collection and data release.

through the ESO archive. They form a simple hierarchical structure where the Phase 3 programme is at the top, followed by data *collection* and data *releases* at the bottom. A data release must be associated to one, and only one, data collection. The hierarchy of the Phase 3 concepts is illustrated in Figure 2. The data collection allows the data from a given programme to be organised according to high-level criteria into self-consistent groups, which the archive user then can browse and access.

As an example of the application of these concepts, a Phase 3 programme is then one of the survey projects (e.g., VHS), one data collection of which may correspond to one target object, or a target region in case of a survey (e.g., the South Galactic Pole for the VHS survey). The collection name must be defined when starting data submission and cannot be revised at a later stage by the user. Each data release must also be supported by a release description that specifies content, data properties and form.

The individual data release may be considered to be a version of the data collection. Subsequent data releases for the same collection may follow the initial release to add more data according to the progress of the observing programme (e.g., Release 2, 3, etc.). The Phase 3 infrastructure allows the data provider to complement, update or supersede a previous release according to the PI's strategy for data publication.

Phase 3 process and tools

The Phase 3 process consists of the following steps that are carried out by the PI/Col and the ESO staff working in the External Data Products (EDP) group. PI/Col activities include the definition of new data collections and releases, data preparation and validation, data upload to the ESO staging area, creation and upload of the release description, and finally closing the release. The PI of an ESO programme can delegate the Phase 3 process to one or more people to distribute the effort of data submission and release preparation. Multiple delegates can, in principle, work on the same data release but it is the sole responsibility of the PI to finally ensure the overall consistency.

On the ESO side, the EDP's activities are to create the Phase 3 programme, validate the submitted data, then carry out their archival and publication. The validation of the submitted data includes the automatic validation of the data against the published format standards for data products, including their header keywords, as a first step, and the verification of the data release content as a second step. During scientific validation, the completeness and consistency of the data release description with respect to the submitted products is examined

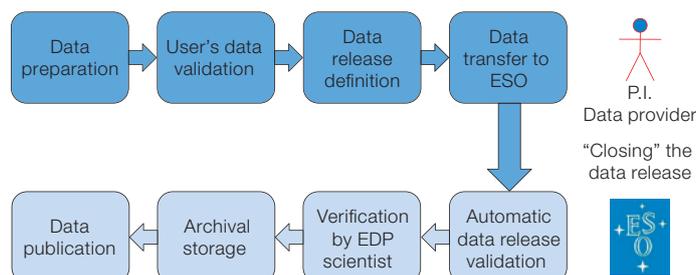


Figure 3. Schematic diagram illustrating the Phase 3 process and responsibilities.

and spot checks of the metadata content and the reported quality parameters are carried out. A summary diagram of the Phase 3 process and the various responsibilities is shown in Figure 3.

Tools

The Phase 3 infrastructure consists of the following components:

- a web application — the Phase 3 release manager — that allows the PI to manage the Phase 3 delegation, and define collections and releases;
- the release validator, which is a command-line application that verifies the data standard and validity of the header keywords against predefined rules;
- the FTP server (phase3ftp.eso.org) that is used in the Phase 3 process by the PI/Col.

On the ESO side, the release manager is accessed in “operator mode” at the beginning of the Phase 3 process, to create a new Phase 3 programme, and at the end, after the data products pass the science validation, for their ingestion into the ESO archive. Applying for the automatic ingestion of the data products extracts the header information and loads this information into the ESO metadata repository, while all data files are sent to the bulk storage system in the ESO archive. Two independent copies of all the data are kept for reasons of safety. A summary showing the different interfaces of the Phase 3 infrastructure with users and the archive facility is displayed in Figure 4.

An important aspect of the support provided by the EDP group is the monitoring of the submitted data products with respect to the executed observations. To this end a dedicated application is used, which accesses the metadata repository and allows the association between data

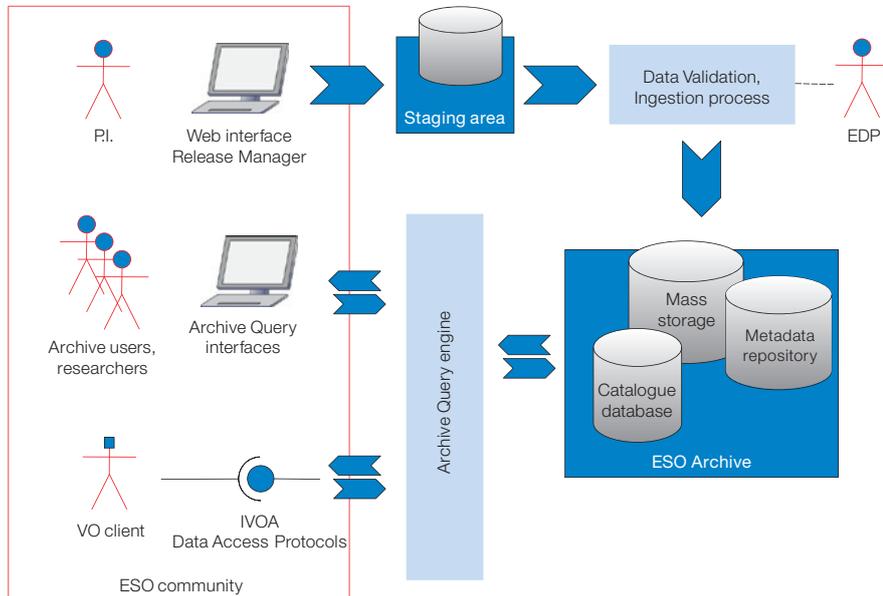


Figure 4. Diagram illustrating the Phase 3 data flow and interfaces with users and the ESO Science Archive Facility. The staging area provides 15 TB of disk space.

product files and the successfully executed OBs at the survey telescopes stored in the Phase 2 database.

Data standards

The definition of data standards and the deployment of the rules as the basis of the automatic validation are key responsibilities of the EDP group. Data standards are required to characterise the level of data reduction and calibration, to track provenance, which allows ESO to monitor survey progress, and finally to support the query for specific data products and VO protocols via the ESO archive interfaces.

For the VISTA data, a number of product types were identified: tiles, pawprints, stripe images, weight maps and source lists. Their definition and relevant keywords are illustrated on the Phase 3 webpage¹.

What's next? Publication of catalogues and internal data products

While supporting the Phase 3 process for ESO programmes and developing new data standards, the EDP group is

engaged in further developing the Phase 3 infrastructure so as to be able to accept catalogues and internal data products (IDPs).

Catalogues

As high-level data products, the resulting source catalogues from the public survey projects represent particularly important results. They are delivered at the milestones set by the major survey releases, the first of which is on 1 October 2011, a year and a half after the beginning of science operation with VISTA. Source catalogues require the data from different bands and tiles to be merged and a global calibration across tiles to be applied. The catalogues are different from the source lists, which are per-tile products and can be downloaded as entire FITS tables. The catalogue contents will be searchable via a dedicated query interface in the ESO archive. Basic functionalities will be supported to allow the archive user to carry out searches by position as well as by non-positional source parameters for sources in any areas of the southern sky, for further scientific selection and investigation on the user's computer.

IDPs

An important aspect of the Phase 3 concepts is that they were developed to support the publication of the data products generated by ESO as part of the quality control process. The DPD fore-

sees that the first IDPs to be ingested and accessible by the community will be those generated by the most advanced pipelines. These IDPs will be made available via the query interfaces in the science archive domain dedicated to data products.

Phase 3 in operation

The Phase 3 tools were developed by the Data Flow Infrastructure Department, and the operational deployment was carried out by the Operations Technical Support Department within ESO. The project manager for the Phase 3 Infrastructure is Remco Slijkhuis, in collaboration with the EDP group.

Inputs from public survey users were collected during a dedicated one-day workshop on VISTA data products at ESO on 30 November 2010. The presentation of the Phase 3 infrastructure and tutorials were followed by a joint discussion with the VISTA survey teams, and whenever possible, the requests by PIs for extra functionality were implemented in the Phase 3 infrastructure that was deployed on 10 March 2011.

The current Phase 3 submission for the VISTA public surveys is ongoing with 6.8 terabytes of data products currently uploaded on the Phase 3 staging area. Once the scientific validation is completed by the EDP group, these data products will be safely stored, and then become available for community access via the archive query interfaces.

For enquiries on the Phase 3 process or to initiate a Phase 3 submission, please contact usd-help@eso.org, quoting Subject: Phase 3.

References

- Arnaboldi, M. et al. 2007, *The Messenger*, 127, 28
- Capaccioli, M., Mancini, D. & Sedmak, G. 2005, *The Messenger*, 120, 10
- Emerson, J., McPherson, A. & Sutherland, W. 2006, *The Messenger*, 126, 41

Links

¹ Phase 3 web information: <http://www.eso.org/sci/observing/phase3.html>